

AMENDMENTS TO THE SPECIFICATION

Please amend the paragraph starting on line 1 of page 49 as follows:

14-3-3 Family (14_3_3; Pfam Pfam Accession No. PF00244). SEQ ID NO:1053 corresponds to a sequence encoding a 14-3-3 protein family member. The 14-3-3 protein family includes a group of closely related acidic homodimeric proteins of about 30 kD first identified as very abundant in mammalian brain tissues and located preferentially in neurons (Aitken et al. *Trends Biochem. Sci.* (1995) 20:95-97; Morrison *Science* (1994) 266:56-57; and Xiao et al. *Nature* (1995) 376:188-191). The 14-3-3 proteins have multiple biological activities, including a key role in signal transduction pathways and the cell cycle. 14-3-3 proteins interact with kinases (e.g., PKC or Raf-1), and can also function as protein-kinase dependent activators of tyrosine and tryptophan hydroxylases. The 14-3-3 protein sequences are extremely well conserved, and include two highly conserved regions: the first is a peptide of 11 residues located in the N-terminal section; the second, a 20 amino acid region located in the C-terminal section. ~~The consensus patterns are as follows: 1) R-N-L-[LIV]-S-[VG]-[GA]-Y-[KN]-N-[IVA]; 2) Y-K-[DE]-S-T-L-I-[IM]-Q-L-[LF]-[RHC]-D-N-[LF]-T-[LS]-W-[TAN]-[SAD].~~

Please amend the paragraph starting on line 32 of page 50 as follows:

(FKH; Pfam Accession No. PF00250). SEQ ID NO:925 corresponds to a gene encoding a polypeptide comprising a forkhead domain. The forkhead domain (also known as a "winged helix") is present in a family of eukaryotic transcription factors, and is a conserved domain of about 100 amino acid residues that is involved in DNA-binding (Weigel et al. *Cell* (1990) 63:455-456; Clark et al. *Nature* (1993) 364:412-420). Mammalian genes that comprise a forkhead domain include those encoding: 1) transcriptional activators (e.g., HNF-3-alpha, -beta, and -gamma proteins, which interact with the cis-acting regulatory regions of a number of liver genes); 2) interleukin-enhancer binding factor (ILF), which binds to purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter and is involved in both positive and negative regulation of important viral and cellular promoter elements; 3) transcription factor BF-1, which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon; 4) human HTLF,

which binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR); 5) transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4; 6) human AFX1 which is involved in a chromosomal translocation that causes acute leukemia; and 7) human FKHR which is involved in a chromosomal translocation that causes rhabdomyosarcoma. The fork domain is highly conserved, and is detected by two consensus patterns: the first corresponding to the N-terminal section of the domain; the second corresponding to a heptapeptide located in the central section of the domain. ~~The consensus patterns are as follows: 1) [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-x(3,4)-[AC]-[LIM]; and 2) W-[QKR]-[NS]-S-[LIV]-R-H.~~

Please amend the paragraph starting on line 19 of page 51 as follows:

Helicases conserved C-terminal domain (helicase_C; Pfam Accession No. PF00271). SEQ ID NOS:227 and 1058 represent polynucleotides encoding novel members of the DEAD/H helicase family. The DEAD box family comprises a number of eukaryotic and prokaryotic proteins involved in ATP-dependent, nucleic-acid unwinding. All DEAD box family members of the above proteins share a number of conserved sequence motifs, some of which are specific to the DEAD family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman, Nature (1988) 333:22 and Nature (1988) 333:578; http://www.expasy.ch/www/linder/HELICASES_TEXT.html). One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins (Wassarman D.A., et al., Nature (1991) 349:463; Harosh I., et al., Nucleic Acids Res. (1991) 19:6331; Koonin E.V., et al., J. Gen. Virol. (1992) 73:989; http://www.expasy.ch/www/linder/HELICASES_TEXT.html). ~~The following signature patterns are used to identify member for both subfamilies: 1) [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN]; and 2) [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR].~~

Please amend the paragraph starting on line 34 of page 51 as follows:

Kazal serine protease inhibitors family signature (Kazal; Pfam Accession No. PF00050). SEQ ID NO:97 corresponds to a polynucleotide of a gene encoding a serine protease inhibitor of the Kazal inhibitor family (Laskowski *et al. Annu. Rev. Biochem.* (1980) 49:593-626). The basic structure of Kazal serine protease inhibitors such a type of inhibitor is described at Pfam Accession No. PF00050. Exemplary proteins known to belong to this family include: pancreatic secretory trypsin inhibitor (PSTI), whose physiological function is to prevent the trypsin-catalyzed premature activation of zymogens within the pancreas; mammalian seminal acrosin inhibitors; canidae and felidae submandibular gland double-headed protease inhibitors, which contain two Kazal-type domains, the first one inhibits trypsin and the second one elastase; a mouse prostatic secretory glycoprotein, induced by androgens, and which exhibits anti-trypsin activity; avian ovomucoids; chicken ovomucoid; and the leech trypsin inhibitor Bdellin B-3. ~~The consensus pattern is as follows: C-x(7)-C-x(6)-Y-x(3)-C-x(2,3)-C, where the four C's are involved in disulfide bonds.~~

Please amend the paragraph starting on line 22 of page 52 as follows:

Neurotransmitter-Gated Ion-Channel (neur_chan; Pfam Accession No. PF00065). SEQ ID NO:1078 corresponds to a sequence encoding a neurotransmitter-gated ion channel. Neurotransmitter-gated ion-channels, which provide the molecular basis for rapid signal transmission at chemical synapses, are post-synaptic oligomeric transmembrane complexes that transiently form a ionic channel upon the binding of a specific neurotransmitter. Five types of neurotransmitter-gated receptors are known: 1) nicotinic acetylcholine receptor (AChR); 2) glycine receptor; 3) gamma-aminobutyric-acid (GABA) receptor; 4) serotonin 5HT3 receptor; and 5) glutamate receptor. All known sequences of subunits from neurotransmitter-gated ion-channels are structurally related, and are composed of a large extracellular glycosylated N-terminal ligand-binding domain, followed by three hydrophobic transmembrane regions that form the ionic channel, followed by an intracellular region of variable length. A fourth hydrophobic region is found at the C-terminal of the sequence. ~~The consensus pattern is: C-x-[LIVMFQ]-x-[LIVMF]-x(2)-[FY]-P-x-D-x(3)-C, where the two C's are linked by a disulfide bond.~~

Please amend the paragraph starting on line 18 of page 53 as follows

Protein phosphatase 2A regulatory subunit PR55 signatures (PR55; Pfam Accession No. PF01240). SEQ ID NO:1028 corresponds to a gene encoding a protine phosphatase 2A reguatory subunit. Protein phosphatase 2A (PP2A) is a serine/threonine phosphatase involved in many aspects of cellular function including the regulation of metabolic enzymes and proteins involved in signal transduction. PP2A is a trimeric enzyme that consists of a core composed of a catalytic subunit associated with a 65 Kd regulatory subunit (PR65), also called subunit A; this complex then associates with a third variable subunit (subunit B), which confers distinct properties to the holoenzyme (Mayer *et al. Trends Cell Biol.* (1994) 4:287-291). One of the forms of the variable subunit is a 55 Kd protein (PR55) which is highly conserved in mammals (where three isoforms are known to exist). This subunit may perform a substrate recognition function or be responsible for targeting the enzyme complex to the appropriate subcellular compartment. ~~Two perfectly conserved sequences of 15 residues, one located the N-terminal region, the other in the center of the protein, serve as the basis for the consensus patterns: 1) E-F-D-Y-L-K-S-L-E-I-E-E-K-I-N; 2) N-[AG]-H-[TA]-Y-H-I-N-S-I-S-[LIVM]-N-S-D~~

Please amend the paragraph starting on line 9 of page 54 as follows:

The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the alignment in (Hanks, *et al., FASEB J.* (1995) 9:576) and covers the entire catalytic domain. ~~The consensus patterns are as follows: 1) [LIV]-G-[P]-G-[P]-[FYWMGSTNH]-[SGA]-[PW]-[LIVCAT]-[PD]-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K, where K binds ATP; 2) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3), where D is an active-site residue; and 3) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC], where D is an active-site residue.~~

Please amend the paragraph starting on line 17 of page 54 as follows:

Ras family proteins (ras; Pfam Accession No. PF00071). SEQ ID NO:527 represents polynucleotides encoding the ras family of small GTP/GDP-binding proteins (Valencia *et al.*, 1991, *Biochemistry* 30:4637-4648). Ras family members generally require a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase

activity. Among ras-related proteins, the highest degree of sequence conservation is found in four regions that are directly involved in guanine nucleotide binding. The first two constitute most of the phosphate and Mg²⁺ binding site (PM site) and are located in the first half of the G-domain. The other two regions are involved in guanosine binding and are located in the C-terminal half of the molecule. Motifs and conserved structural features of the ras-related proteins are described in Valencia et al., 1991, *Biochemistry* 30:4637-4648. ~~A major consensus pattern of ras proteins is: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y.~~

Please amend the paragraph starting on line 12 of page 56 as follows:

Zinc Finger, C2H2 Type (Zincfing_C2H2; Pfam Accession No. PF00096). Several sequences corresponded to polynucleotides encoding members of the C2H2 type zinc finger protein family, which contain zinc finger domains that facilitate nucleic acid binding (Klug et al., *Trends Biochem. Sci.* (1987) 12:464; Evans et al., *Cell* (1988) 52:1; Payre et al., *FEBS Lett.* (1988) 234:245; Miller et al., *EMBO J.* (1985) 4:1609; and Berg, *Proc. Natl. Acad. Sci. USA* (1988) 85:99). In addition to the conserved zinc ligand residues, a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld et al., *J. Biomol. Struct. Dyn.* (1993) 11:557) The best conserved position, which is generally an aromatic or aliphatic residue, is located four residues after the second cysteine. ~~The consensus pattern for C2H2 zinc fingers is: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H. The two C's and two H's are zinc ligands.~~

Please amend the paragraph starting on line 23 of page 56 as follows:

Zinc finger, C3HC4 type (RING finger), signature (Zincfing_C3H4; Pfam Accession No. PF00097). SEQ ID NOS:805 and 1078 represent polynucleotides encoding a polypeptide having a C3HC4 type zinc finger signature. A number of eukaryotic and viral proteins contain this signature, which is primarily a conserved cysteine-rich domain of 40 to 60 residues (Borden K.L.B., et al., *Curr. Opin. Struct. Biol.* (1996) 6:395) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D structure of the zinc ligation system is unique to the RING domain and is referred to as the "cross-brace" motif. ~~The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C. The signature pattern for the~~

~~C3HC4 finger is based on the central region of the domain: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA].~~

Please amend the paragraph starting on line 33 of page 56 as follows:

Zinc finger, CCHC type (Zincfing_CCHC; Pfam Accession No. PF00098). SEQ ID NOS:693,973, and 1078 correspond to genes encoding a member of the family of CCHC zinc fingers. Because the prototype CCHC type zinc finger structure is from an HIV protein, this domain is also referred to as a retroviral-type zinc finger domain. The family also contains proteins involved in eukaryotic gene regulation, such as *C. elegans* GLH-1. The structure is an 18-residue zinc finger; no examples of indels in the alignment. ~~The motif that defines a CCHC type zinc finger domain is: -C-X2-C-X4-H-X4-C (Summers J Cell Biochem 1991 Jan;45(1):41-8).~~ The domain is found in, for example, HIV-1 nucleocapsid protein, Moloney murine leukemia virus nucleocapsid protein NCp10 (De Rocquigny *et al. Nucleic Acids Res.* (1993) 21:823-9), and myelin transcription factor 1 (Myt1) (Kim *et al. J. Neurosci. Res.* (1997) 50:272-90).

Please amend the paragraph starting on line 9 of page 65 as follows:

In addition, pools of selected clones, as well as libraries containing specific clones, were assigned an "ES" number (internal reference) and deposited with the ATCC. Table 21 below provides the ATCC Accession Nos. of the ES deposits, all of which were deposited on or before May 13, 1999. ~~The names of the clones contained within each of these deposits are provided in the tables numbered 22 and greater (inserted before the claims).~~